

Group behavior modeling based on: stochastic cellular automata with variable memory, systems of differential kinetic equations and analysis of social processes' time series

Zhukov D.

Russian Technological University
Moscow, Russia
zhukovdm@yandex.ru

Smychkova A.

Russian Technological University
Moscow, Russia
nsmych@yandex.com

Abstract — The paper considers a set of models for the description, analysis, and modeling of group behavior by user actions in complex social systems. In particular, electoral processes in which preferences are selected from several possible ones can be considered. The dynamics of changes in the preferences of voters can be described graphically diagram of possible transitions between States, on the basis of which it is possible to obtain a system of differential kinetic equations describing the process. The analysis of decisions shows the possibility of existence within the model, different modes of change of preferences of voters. In addition, the paper describes the developed model of stochastic cellular automata with variable memory, in which at each step of the interaction process between its cells a new network of random links is established, the minimum and the maximum number of which is selected from a given range. At the initial time, a cell of each type is assigned a numeric parameter that specifies the number of steps during which it will retain its type (cell memory). The transition of cells between States is determined by the total number of cells of different types with which there was interaction at a given number of memory steps. After the number of steps equal to the depth of memory, its transition to the type that had the maximum value of its amount occurs. The effect of external factors (such as media) on changes in node types can be set for each step using a transition probability matrix. Processing of sociological data of the electoral campaign of 2015-2016 at the choice of the USA President using the method of almost – periodic functions allowed to estimate the parameters of the complex of created models and use them to describe, analyze and model the group behavior of voters. The studies show a good correspondence between the data observed in sociology and calculations using a set of developed models. Some sets of coefficients in differential equations and cellular automata models show oscillating and almost - periodic changes in the preferences of voters, which largely coincides with real observations.

Keywords — *group behavior, cellular automata, stochastic cellular automata, differential kinetic equations, social process.*

I. INTRODUCTION

The analysis of changes and modeling of dynamics of processes of group behavior is of great interest and is very relevant. In particular, the events of 2016: Brexit UK and USA elections showed a weak consistency of modern sociological models of forecasting the results of group choice. Most polling

agencies predicted the victory of Hillary Clinton and the failure of the UK from leaving the EU. However, even based on statistically significant sociological samples in a survey of respondents about their intentions in voting, most services made incorrect predictions. On the one hand, this was due to the high volatility of the data obtained in the surveys, and on the other hand, the fact that the existing methods of modeling and forecasting are not effective.

In our opinion, it is interesting to develop a set of models based on the use of micro-kinetic and macro-kinetic approaches. In the micro-kinetic approach, it is necessary to consider the interaction at the level of individual users in the description, analysis, and management of group behavior. In the macro-kinetic description of the processes it is necessary to consider the dynamics of the integrative parameters describing the system as a whole.

It should be noted that in both micro-kinetic and macro-kinetic approaches, a set of parameters is needed to create models, the specific values of which and the relations between which can be determined based on the analysis of the time series of social processes of group behavior. Very often, the values of the parameters that describe the processes of group behavior in social, economic and socio-technical systems have oscillatory and trend components.

For effective analysis of fluctuations, it is necessary to separate the trend and fluctuations without losing significant information about the process. In the analysis of social processes, the use of the method of almost – periodic functions is more preferable than traditional methods of harmonic analysis based on Fourier transforms. This is due to a number of reasons, the Main of which is the presence in social processes of the human factor, which plays a significant role, which leads to fuzziness, uncertainty and non-deterministic characteristics of processes. While Fourier analysis methods are based on the fact that the signals have a harmonic nature.

In our opinion, the macro-kinetic model can be constructed on the basis of consideration of group behavior by means of a method of graphic diagrams of transitions between States of all considered system as a whole. Further, based on the consideration of the obtained diagram, a system of differential

kinetic equations can be written, the solution of which, with a certain set of coefficients, will simulate group behavior.

To create a micro-kinetic model, an approach developed on the basis of stochastic cellular automata with variable memory can be used.

II. TIME SERIES ANALYSIS OF GROUP BEHAVIOR'S COMPLEX PROCESSES IN SOCIAL SYSTEMS USING THE METHOD OF ALMOST PERIODIC FUNCTIONS

As an example for data processing using the method of almost – periodic functions, the results of observations of voters' preferences conducted in the United States for 500 days, from July 1, 2015, to November 7, 2016, during the presidential campaign of 2016 (data taken from the resource: http://www.realclearpolitics.com/epolls/2016/president/us/general_election_trump_vs_clinton-5491.html#polls). The data presented in figure 1 have a trend (down and up baselines) and oscillatory components.

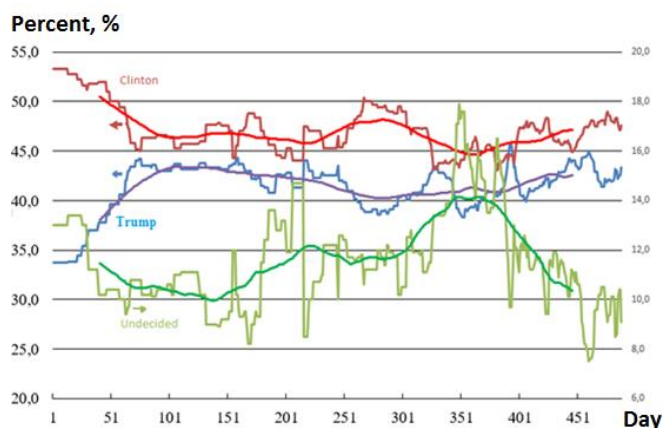


Fig. 1. Initial and smoothed data on the preferences of USA voters during the presidential campaign of 2016 (Trump, Clinton and undecided voters)

We define the characteristics of the fluctuations in the preferences of voters. Trend elimination and determination of periodic components is achieved by converting the original series into a new series according to the formula [1]:

$$a(t_k) = \ln \frac{y(t_k - \Delta t_m) \cdot y(t_k + \Delta t_m)}{y(t_k)^2} \quad (1)$$

where $y(t_k - \Delta t_m)$, $y(t_k)$, $y(t_k + \Delta t_m)$ - are the values at the appropriate time moments, t_k - the moments of registration of the measured characteristics, Δt_m - a fixed trial time interval. The result of the transformation is a series $a(t_k)$ with a value of expectation close to zero.

With a trial shift equal to zero, the initial series is a trend, with an increase in the value of the trial shift, the trend degenerates into a straight line, with a significant decrease in the number of elements of the series. The use of shift functions based on the metrics of functional analysis and the theory of almost - periodic functions makes it possible to effectively determine the values close to the periods (almost -

where n - the number of original series' points, τ_k - test slide shift.

It should be noted that for different trial shifts of time intervals, due to the specifics of the data (a large number of consecutive values without changes), finding the values of almost - periods will vary in accuracy. It is necessary to choose such a test shift, in which the amplitude of the oscillations of the obtained series will be the smallest.

Due to the fact that the trial shift will characterize the range in which the alignment will occur relative to the trend, you should carefully select this value. If the values of the test slide are too small, there will be a large number of oscillations, relative to zero, by small values. As a result, the shift function will show a larger number of lows, respectively, to determine almost the period will be difficult due to the ambiguity of the results. Thus, the choice of a sufficiently large trial changes, the effect of reducing cardinality of a row, the minimum definition of a function would be impossible (the number will be rising relative to its trend).

The analysis of the observed electoral data in the USA presidential campaign of 2016 using the method of almost – periodic functions showed that Donald Trump had one almost-period in the fluctuations of preferences of voters equal to 80 days, which, taking into account the error of determination, practically remained for the entire campaign. Hillary Clinton had two days-periods: 50 and 130 [2].

The obtained data can be used to create models of analysis, description, and control of group processes, based on stochastic cellular automata with variable memory [3] and systems of kinetic differential equations [4], in determining the parameters of these models and the relations between them.

III. GROUP BEHAVIOR'S MODEL OF USERS IN COMPLEX SOCIAL SYSTEMS BASED ON SYSTEMS OF DIFFERENTIAL KINETIC EQUATIONS

Graphical charts can be used to describe group behavior and select preferences. Let's denote the number of voters who are at the time t in the state in which they prefer this candidate (let's call him "A") as $y_1(t)$. The number of voters in a state in which they will vote for candidate "B" at a given time t will be designated as $y_2(t)$. The number of voters in a neutral state (not determined for any of the candidates), denote as $y_3(t)$. The objective of the model is to determine the dependence of changes over time, the number of voters who will give preference to each of the candidates in the elections.

Under the influence of the media and PR companies of candidate "A", the views of candidate "B" voters may change over time, so that they may first move to a neutral (candidate "C") state $y_3(t)$, and then begin to support candidate "A". Similarly, the same can happen to supporters of candidate "A". We introduce the following parameters of the model: τ_1 - time of change (or change) of preferences of voters' $y_1(t)$ candidate "A", that is, these voters become vulnerable to new attempts of supporters' $y_2(t)$ to convince them to support this candidate after a while τ_1 (change of mood); τ_2 – time of change of views of voters $y_2(t)$. The values of times τ_1 and τ_2 depend on the action of the media and PR companies of candidates. For example, the more effectively the candidate "B" is campaigning, the less time τ_1 will be the change of views of

This research was supported by Russian Foundation for Basic Research (RFBR), grant No. 16-29-09458.

periods). A special case of such functions is:

$$\chi(\tau_k) = \frac{1}{n-k} \sum_{j=1}^{n-k} |a(t_j + \tau_k) - a(t_j)| \quad (2)$$

candidate “a” voters and the more time τ_2 – change of views for candidate “B” voters.

With a larger number of candidates, the graph shown in figure 2 will contain a larger number of nodes and links (with m – candidates, the number of interconnections will be $m(m-1)/2$). Note that the node responsible for mass media will have unidirectional links with other nodes.

As an example, consider the chart of transitions between the preferences of voters in relation to candidates (see Fig. 2) during campaign for elections of the President of the United States in 2016.

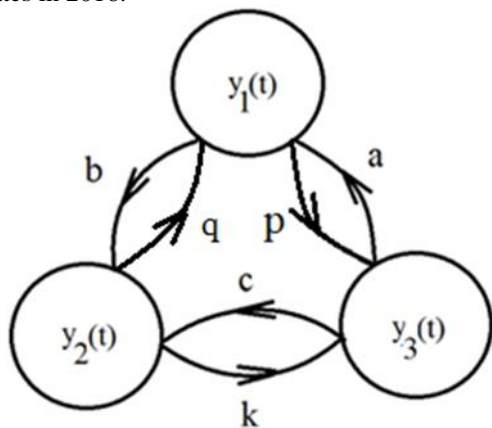


Fig. 2. Processes diagram of changing the preferences of voters during elections

Kinetic differential equations (3) – (5) describes the change of mood over time can be written on the basis of the transitions of a graphical diagram between the preferences of voters.

$$\frac{dy_1(t)}{dt} = ay_1(t - \tau_1)y_3(t) - py_1(t - \tau_1) - by_1(t - \tau_1)y_2(t - \tau_2) + qy_1(t - \tau_1)y_2(t - \tau_2) = ay_1(t - \tau_1)y_3(t) - py_1(t - \tau_1) - (b - q)y_1(t - \tau_1)y_2(t - \tau_2) \quad (3)$$

$$\frac{dy_2(t)}{dt} = cy_2(t - \tau_2)y_3(t) - ky_2(t - \tau_2) + by_1(t - \tau_1)y_2(t - \tau_2) - qy_1(t - \tau_1)y_2(t - \tau_2) = cy_2(t - \tau_2)y_3(t) - ky_2(t - \tau_2) + (b - q)y_1(t - \tau_1)y_2(t - \tau_2) \quad (4)$$

$$\frac{dy_3(t)}{dt} = -ay_1(t - \tau_1)y_3(t) - cy_2(t - \tau_2)y_3(t) + ky_2(t - \tau_2) + py_1 \quad (5)$$

where a, b, c, p, q and k – coefficients of differential equations, the ratio between which and their value is determined on the basis of nonlinear analysis of the time series of the processes under consideration, for example, on the basis of the method of almost-periodic functions.

To explain the model, consider in more detail one of the kinetic equations, for example, (4). Member of equation $\frac{dy_2(t)}{dt}$ describes the rate of change of the number of voters supporting candidate “B”; $cy_2(t - \tau_2)y_3(t)$ determines the amount of these constituents due to the influence of supporters $y_2(t - \tau_2)$ candidate “B” on undecided voters $y_3(t)$; $ky_2(t - \tau_2)$ – defines the decrease of the supporters of candidate “B” due to the obsolescence of their views τ_2 and this process has a linear character; $by_1(t - \tau_1)y_2(t - \tau_2)$ determines the increase due to agitation of voters $y_1(t - \tau_1)$ favoring candidate “A” supporters of candidate “B” $y_2(t - \tau_2)$ (since this process is associated with mutual influence of one another, it is written as a product of functions

$y_1(t - \tau_1)$ and $y_2(t - \tau_2)$; $qy_1(t - \tau_1)y_2(t - \tau_2)$ determines the loss due to the agitation of voters preferring candidate B supporters of candidate “A”. Similarly, the semantic value of the terms of the kinetic equations (3) and (5) is determined.

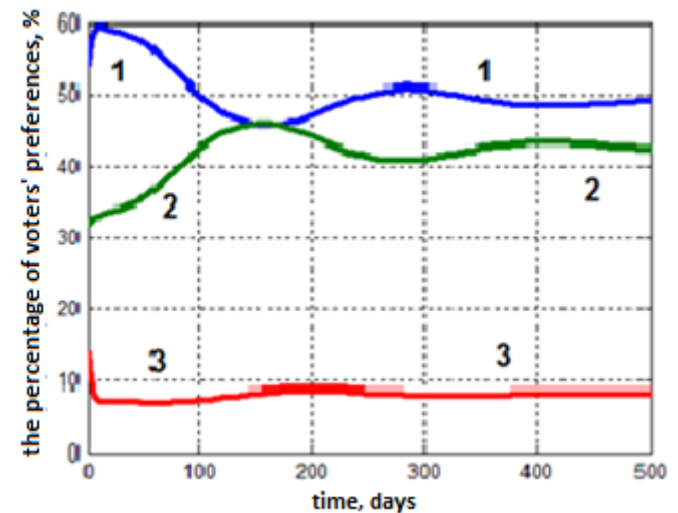
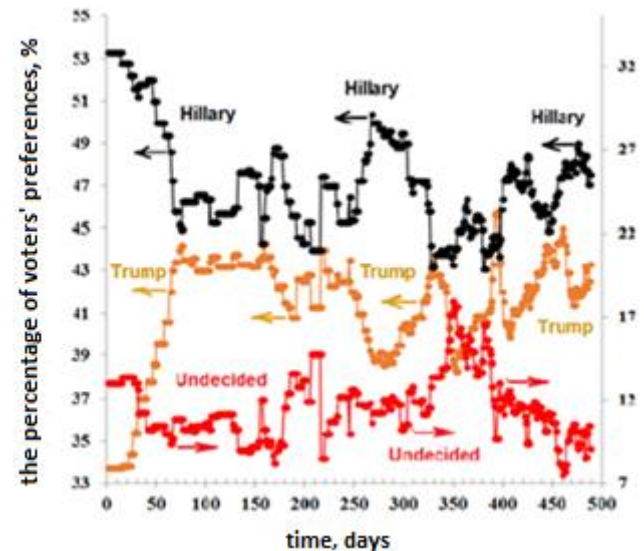


Fig. 3. Comparison of the observed preferences of USA voters during the presidential campaign of 2016 (trump, Hillary and undecided voters) and the results of modeling

The results of the simulations' evaluation with the values of the model parameters: $a=0.0054$; $b=0.00056$; $c=0.0008$; $p=0.024$; $k=0.032$; $q=0.000045$ and values of the times, the attitudes of $\tau_1=43$ and $\tau_2=43$ days are shown in figure 3 (curve 1 – the preferences of the voters for Clinton, curve 2 – for trump, curve 3 – undecided voters). Note that the duration of the change of views of voters for Hillary Clinton $\tau_1=43$ days and $\tau_2=43$ for Donald trump, taken equal to the minimum of the almost – periods of candidates (80-90 days for trump and 50 days for Clinton).

The selection of coefficients in the kinetic differential equations allows to obtain a sufficiently good agreement between the theoretical results of the model and the observed data (see Fig. 3).

IV. GROUP BEHAVIOR MODEL BASED ON STOCHASTIC CELLULAR AUTOMATA WITH VARIABLE MEMORY

In the proposed model of stochastic cellular automata with variable memory at each step of the process between its cells, a new network of random links is established, the minimum and the maximum number of which is selected from a given range. At time $t=0$, the node of each type is assigned a numeric parameter that specifies the number of steps during which it will store its type (cell memory). The transition of nodes between States is determined by the total number of nodes of different types with which there was interaction at a given number of memory steps. After the number of steps equal to the depth of memory, its transition to the type that had the maximum value of its amount occurs. The effect of external factors (such as media) on changes in node types can be set for each step using a transition probability matrix.

When constructing the proposed model, we begin by specifying the types and the initial number of different cells, each of which can be set in accordance with certain properties of the cells of the cellular automaton. When cells interact or are externally exposed to them, their properties can change in accordance with the specified rules (one type of cell passes to another). When modeling processes in social systems, such as election campaigns, the properties of cells can be, for example, the preferences of voters for a particular candidate (or political party). Note that in other processes cells may have other properties, so to build a model and formalize the description, we introduce the concept of "color". Thus, each type of cells will have its own "color", which can change during the operation of the cellular automaton. The initial number of cells of each type is set before starting the cellular automaton at time $t=0$.

At each step of the process, a new network of random links between nodes in the system is constructed, rather than considering the physical movement of cells. In this case, the minimum and the maximum number of links of each node is selected from a given range (which causes the construction of a stochastic network at each step of the process) modeling.

Consider the dynamics of changes in the state of the cellular automaton simulating the election campaign Trump-Clinton. The average number of connections of any cell of the cellular automaton at each step of the process, as an example, take from 3 to 7, this is approximately the number of discussions of political topics in one month of one, given by a person with others. Processing of sociological data using the method of almost – periodic functions shows that the value of fluctuations in the mood of trump voters and undecided was 86 days, and Hillary Clinton voters 50, so the depth of memory of Hillary Clinton voters can be taken equal to one conditional step, and trump voters and undecided two steps (because almost – Clinton period is less). The duration of one step will be 50 days, and the duration of the entire 500-day election campaign will be 10 modeling steps. As one of the possible options for modeling, the values of probabilities of transitions between the States of cells under the influence of external factors, presented in table 1, were chosen.

TABLE I. THE PROBABILITIES OF TRANSITIONS IN ONE STEP UNDER THE ACTION OF RANDOM EXTERNAL FACTORS.

	Clinton	Trump	Undecided
Clinton	0,790	0,030	0,180
Trump	0,025	0,775	0,200
Undecided	0,272	0,235	0,493

Simulation results (see Fig. 4) show that by the end of the election campaign, the values of voters' preferences almost converge to one point in the area of 38 %, which is in good agreement with the observed values of voters' preferences in practice. However, it should be noted that the results are more qualitative than quantitative.

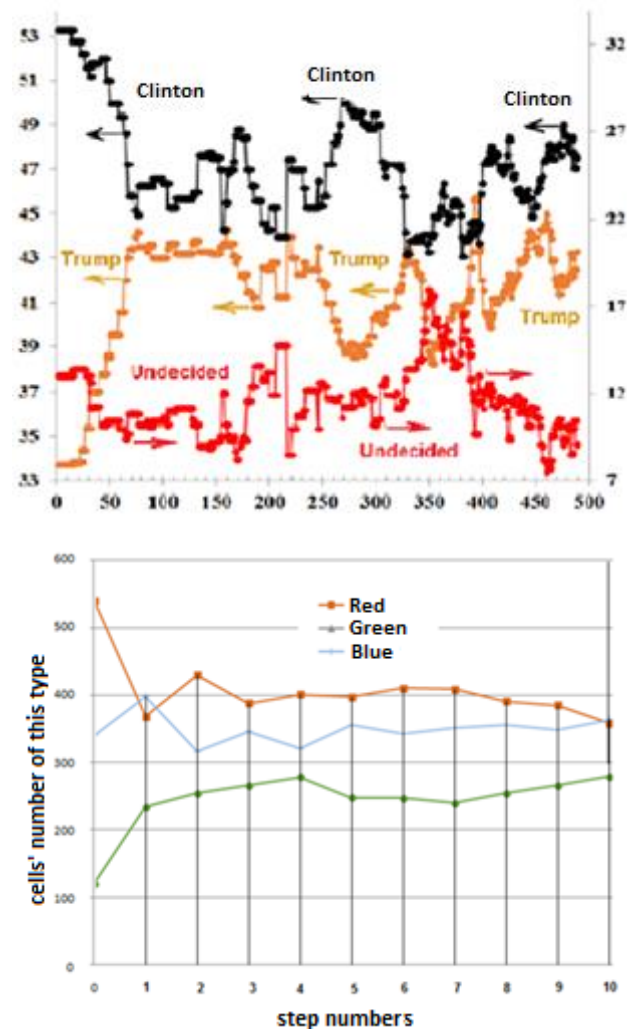


Fig. 4. The simulation of transitions of a cellular automaton describing the election campaign of Clinton - Trump

V. SUMMARY AND CONCLUSIONS

The methods of time series analysis describing complex social processes on the basis of the method of almost periodic functions are developed. In the analysis of social, including electoral processes, the use of methods based on Fourier transforms is not effective, since their application is based on the fact that the formation of oscillations is a superposition of harmonic oscillations. However, the real processes of changing the preferences of voters, according to the mechanisms of their formation, may not meet this condition,

and are nonlinear, due to the presence of human factors leading to fuzzy, uncertain and non-deterministic characteristics of the processes. The obtained characteristics can be used as parameters in macro and microkinetic models.

Macrokinetic (based on a system of differential equations describing transitions between States) and microkinetic (based on stochastic cellular equations with variable memory) models for describing group behavior in social systems are developed. Comparison of the results of the models' behavior with the observed sociological data shows that: they can be used to describe group behavior in social systems, for example, group choice during election campaigns. These models are not contradictory and can complement each other in predicting election results.

Acknowledgment

This research was supported by Russian Foundation for Basic Research (RFBR), grant No. 16-29-09458.

References

- [1] B. M. Levitan. Almost – periodic functions. M., 1953. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
- [2] Dmitry Zhukov, Tatiana Khvatova. The 2016 USA election campaign: analysing trends using the method of almost-periodic functions. Proceeding of the 5 th European Conference on Social Media (ECSM 2018), pp. 353- 363.
- [3] Sigov, A.S., Obukhova, A.G., Alyoshkin, A.S., Zhukov, D.O. Mathematical solutions of software for predicting group behavior in a social system based on stochastic cellular automata with memory. CEUR Workshop. Proceedings 2nd International Scientific Conference "Convergent Cognitive Information Technologies", Convergent 2017; 24 -26 November 2017, Volume 2064, 2017, pp. 175-186.
- [4] Zhukov D.O., Alyoshkin A.S., Obukhova A.G. Modelling to be based on Systems of Differential Kinetic Equations to Processes Group Selection Voters during the Electoral Campaign of Trump-Clinton 2015 – 2016. Proceeding The 7th International Conference on Information Communication and Management ICIM'17, August 28-30, 2017, ACM ISBN 978-1-4503-5279-6/17/08, p. 88-94.